

CLAIMS

What is claimed is:

- 1 1. A method for managing a memory system having a plurality of subsystems,
2 comprising the steps of:
3 upon accessing the memory system for a piece of data used by a first
4 process
5 determining the access time to acquire the piece of data in the
6 memory system;
7 comparing the determined access time to a threshold; and
8 taking actions based on the results of the comparing step.
- 1 2. The method of claim 1 further comprising the step of postponing executing the
2 first process and allowing executing a second process, if the step of comparing
3 indicates that the determined access time is close to, equal to, or greater than the
4 threshold.
- 1 3. The method of claim 2 wherein an intelligence performing the steps of postponing
2 and allowing upon a latency manager notifying the intelligence that the determined
3 access time is close to, equal to, or greater than the threshold; the latency manger
4 performing the step of determining independent from the intelligence.
- 1 4. The method of claim 3 wherein the intelligence is selected from a group consisting
2 of a processor working with the memory system, an operating system working
3 with the memory system, software running on the processor, and a memory
4 manager managing the memory system.

- 1 5. The method of claim 1, if the step of comparing indicates that the determined
2 access time is close to, equal to, or greater than the threshold, further comprising
3 the step of monitoring the memory system or a system using the memory system.
- 1 6. The method of claim 1 wherein the determined access time is selected as the
2 longest access time of a plurality of access times each of which corresponds to a
3 memory access in a multiple memory access.
- 1 7. The method of claim 1 further comprising the step of accessing the piece of data in
2 more than one subsystem at the same time; one subsystem having a shorter access
3 time and one subsystem having a longer access time; the determined access time
4 being that of the subsystem having the shorter access time, and, if the piece of data
5 is missed in the subsystem having the shorter access time, then the determined
6 access time being that of the subsystem having the longer access time.
- 1 8. The method of claim 1 further comprising the step of updating a previous
2 determined access time to the determined access time if the determined access
3 time is greater than the previous determined access time.
- 1 9. The method of claim 1 further comprising the step of notifying an intelligence
2 working with the memory system; the intelligence being selected from a group
3 consisting of a processor, an operating system, software running on the processor,
4 and a memory manager managing the memory system; the intelligence performing
5 the step of taking actions.
- 1 10. The method of claim 1 further comprising the step of changing the determined
2 access time upon performing a task selected from a group consisting of changing
3 the threshold, initiating an interrupt to an intelligence working with the memory

4 system, and postponing executing the first process and allowing executing a
5 second process.

1 11. The method of claim 1 wherein the determined access time is selected from the
2 time to access at least one subsystem.

1 12. The method of claim 1 wherein a latency manager performing the step of
2 determining; the latency manager being on the data path between a processor
3 working with the memory system and the plurality of subsystems.

1 13. The method of claim 1 wherein the data is accessed from a subsystem having a
2 shorter access time to a subsystem having a longer access time or in a non-
3 sequential order.

1 14. A method for managing a memory system having a plurality of subsystems,
2 comprising the steps of:
3 earmarking a subsystem;
4 from the plurality of subsystems, determining an order for data to be
5 accessed from a subsystem having a shorter access time to a
6 subsystem having a longer access time; and
7 upon accessing the memory system for a piece of data used by a first
8 process, if the data is missed in the earmarked subsystem, then
9 postponing executing the first process and allowing executing a
10 second process.

1 15. The method of claim 14 wherein an intelligence performing the steps of
2 postponing and allowing upon a latency manager notifying the intelligence that the
3 determined access time is close to, equal to, or greater than the threshold; the

4 intelligence being selected from a group consisting of a processor working with
5 the memory system, an operating system working with the memory system,
6 software running on the processor, a memory manager managing the memory
7 system; the latency manger being part of managing the memory system.

1 16. An apparatus for managing a memory system having a plurality of subsystems,
2 comprising:
3 means for, upon accessing the memory system for a piece of data used by a
4 first process,
5 determining the access time to acquire the piece of data in the
6 memory system;
7 comparing the determined access time to a threshold; and
8 taking actions based on the results of the comparing step.

1 17. The apparatus of claim 16 further comprising means for postponing executing the
2 first process and allowing executing a second process, if the step of comparing
3 indicates that the determined access time is close to, equal to, or greater than the
4 threshold.

1 18. The apparatus of claim 16 wherein the determined access time is selected as the
2 longest access time of a plurality of access times each of which corresponds to a
3 memory access in a multiple memory access.

1 19. The apparatus of claim 16 further comprising means for accessing the piece of data
2 in more than one subsystem at the same time; one subsystem having a shorter
3 access time and one subsystem having a longer access time; the determined access
4 time being that of the subsystem having the shorter access time, and, if the piece of

5 data is missed in the subsystem having the shorter access time, then the determined
6 access time being that of the subsystem having the longer access time.

1 20. An apparatus for managing a memory system having a plurality of subsystems,
2 comprising:

3 means for earmarking a subsystem;

4 means for determining, from the plurality of subsystems, an order for data
5 to be accessed from a subsystem having a shorter access time to a
6 subsystem having a longer access time; and

7 upon accessing the memory system for a piece of data used by a first
8 process, if the data is missed in the earmarked subsystem, then
9 means for postponing executing the first process and allowing
10 executing a second process.

1 21. A computer-readable medium embodying instructions for a computer to perform a
2 method for managing a memory system having a plurality of subsystems, the
3 method comprising the steps of:

4 upon accessing the memory system for a piece of data used by a first
5 process,

6 determining the access time to acquire the piece of data in the
7 memory system;

8 comparing the determined access time to a threshold; and

9 taking actions based on the results of the comparing step.

1 22. The computer-readable medium of claim 21 wherein the method further comprises
2 the step of postponing executing the first process and allowing executing a second
3 process, if the step of comparing indicates that the determined access time is close
4 to, equal to, or greater the threshold.

1 23. The computer-readable medium of claim 21 wherein the determined access time is
2 selected as the longest access time of a plurality of access times each of which
3 corresponds to a memory access in a multiple memory access.

1 24. The computer-readable medium of claim 21 wherein the method further
2 comprising the step of accessing the piece of data in more than one subsystem at
3 the same time; one subsystem having a shorter access time and one subsystem
4 having a longer access time; the determined access time being that of the
5 subsystem having the shorter access time, and, if the piece of data is missed in the
6 subsystem having the shorter access time, then the determined access time being
7 that of the subsystem having the longer access time.

1 25. A computer-readable medium embodying instructions for a computer to perform a
2 method for managing a memory system having a plurality of subsystems, the
3 method comprising the steps of:
4 earmarking a subsystem;
5 from the plurality of subsystems, determining an order for data to be
6 accessed from a subsystem having a shorter access time to a
7 subsystem having a longer access time; and
8 upon accessing the memory system for a piece of data used by a first
9 process, if the data is missed in the earmarked subsystem, then
10 postponing executing the first process and allowing executing a
11 second process.